ON THE DESIGN AND ANALYSIS OF REPEATED MEASURES EXPERIMENTS

by

Walter T. Federer and Ronald P. Kershner

BU-681-M*

July 1979
(revised June 1980)

1. INTRODUCTION AND DEFINITIONS

A commonly occurring phenomenon in experimental investigations is the repetition of treatments on the same sampling unit. When an animal (or human) is used in a nutrition experiment, several different diets in a specified sequence for a specified period of time are given to the animal (or human). The sequence of treatments varies from animal to animal. When a plot of perennial plants is used in a fertilizer or spraying experiment, several different fertilizers of spraying treatments in a specified sequence for a specified period of time are applied to the plot; the sequence of treatments varies for the different plots. Many other types of examples are available. This procedure is sometimes called the repeated measurements situation. This terminology is somewhat inadequate in that it does not indicate that the treatments will change and neither does it indicate that the same sampling unit is used repeatedly. However, we shall conform to current usage and will discuss a number of situations involving a repeated measurements design and the need for flexibility in both the design and analysis. First, a number of definitions are needed.

A population consists of members which are designated as sampling units. A simple random sample of s sampling units from the population is obtained. A treatment is an entity of interest to the investigator. A treatment design is the the selection of the v treatments for the experiment to be conducted. An experimental

---

* No. BU-681-M in the Mimeo Series of the Biometrics Unit, Cornell University.

unit is the smallest unit to which one treatment is applied. The number of experimental units for a specified treatment equals the number of replicates for that treatment. The experiment design is the arrangement of the treatments in the experiment; it is the experimental plan. Since every experiment has a plan of procedure, every experiment has some kind of experiment design. Note that experiment designs in statistics textbooks and in statistics class notes represent but a small fraction of those used in practice.

When a sampling unit (s.u.) is selected from the population and when several treatments are applied to the s.u. in sequence (i.e., there are several experimental units per s.u.), various types of treatment effects are possible. First, there is a direct effect of the treatment observed at the time or in the period when that treatment is applied. Secondly, there may be a residual effect of the treatment for one, two, three, etc. periods beyond the period in which the treatment was applied to the experimental unit (e.u.). Treatment effects lasting for one extra period will be denoted as one-period residual effects; those lasting two periods beyond the period of application will be denoted as two-period residual effects; etc. Also, there are treatments which exhibit effects in all subsequent periods; this effect is designated as a continuing effect of the treatment. The sum of the direct plus residual effects is known as the permanent effect of a treatment. The existence of several kinds of treatment effects needs consideration in both the design and analysis of an experiment. A detailed account of repeated measures designs and analyses for two treatments is given in a paper by Kershner and Federer [1980].

Experiment designs have been constructed such that direct and one-period residual treatment effects can be estimated when a linear model is appropriate. Textbook coverage of this type of experiment design and analysis appears to be

limited to few books, i.e., Cox [1958], Cochran and Cox [1957], Federer [1955], Finney [1960], John, et al. [1972], Kempthorne [1952], and Quenouille [1953]. Federer and Balaam [1972] prepared a bibliography on these designs and their analysis through 1967. Hedayat and Afsarinejad [1975] present a discussion on the construction of some of these designs along with a bibliography. The coverage in the foregoing is for standard situations which do not take into account variations often necessitated by experimental conditions or goals of the experimenter. Some of these variations are discussed herein. In the next section, we use an example to demonstrate the need for flexibility in the statistical design and analysis of data from a repeated measures design on digestibility of foods from diets composed of various kinds of fiber. In section three, a number of variations of textbook repeated measures designs are described.

2. AN EXAMPLE

This example arose in connection with a study of the influence of non-digestible components of fruits, vegetables, grains, and other foods on bowel movement and cancer of the colon in humans. There was considerable speculation, but no quantification of the influence. The purpose of the study was to quantify the influence of diet on bowel movement and microbial flora.

2.1. Population, Sampling Unit, and Sampling Procedure

The population to be studied consisted of 20-30 year old males who were off drugs and antibiotics, nonsmokers, and meat eaters, who had no history of gastro-intestinal disorders or ulcers, and who could and would participate in the study. Diets were to be strictly presecribed over a period of about 80 days; the subjects were to offer daily stool and occasional blood samples and be willing to accept payment. Such restrictions greatly reduced the size of this sub-population as

compared to all males in this age bracket. The sampling unit then is a male person. The sampling procedure was to select a simple random sample from among those showing up at meetings and who met the prescribed qualifications.

2.2 Treatment Design

A basal diet, E, supplied approximately 2400 calories per day; this was the estimated energy need for an average subject. Of this amount about 250 calories were provided by white bread. This diet was nearly fiber free with only potatoes and bread providing small amounts of fiber. For diet A, the bread was baked with coarse bran (32 grams per subject); for diet B, the bread was prepared with fine bran (32 grams per subject); for diet C, the bread was baked with cellulose (12 grams per subject) to provide the bulk; and for diet D, the bread was prepared with extracted cabbage (17 grams per subject). Calcium and iron supplements were added to bring the mineral contents of all diets to a common level. Consumption of coffee, tea, and wine (on Saturday evening only) was optional, but constant intakes at standard times were maintained if the subject chose to have beverage with his diet. Deionized water was available to satisfy liquid requirements.

2.3. Experiment Design and Experiment Unit

There was a time constraint on performing the experiment, since it had to be started before the beginning of the Spring Term and had to end before the Spring vacation period. The experimental unit was to be one diet consumed for a period of perhaps 25 days. It was not known if there would be a carry-over effect of the previous diet or if the period was long enough to observe the infludnce of a diet. It was decided that a sample of 24 males would be obtained and that three treatment periods, i.e., three experimental units per sampling unit, would probably be used. In preparing the design it was decided to add two additional periods to illustrate

what could be done if two additional periods were possible. Also, since it was not known what the length of the period would be, two designs were suggested for use on each of two sets of 12 males. The first experiment design (design 1) is constructed from the three orthogonal latin squares of order four as follows (see, e.g., Federer [1955]):

	male											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	C	D	A	B	B	A	D	C	D	C	B	A
4	D	C	B	A	C	D	A	B	B	A	D	C
5	A	B	C	D	A	B	C	D	A	B	C	D
⋮	⋮											

The 12 students were randomly assigned the numbers 1,2,⋯,12. The second experiment design (design 2) suggested was:

	male											
Period	13	14	15	16	17	18	19	20	21	22	23	24
1	A	A	A	B	B	B	C	C	C	D	D	D
2	A	A	A	B	B	B	C	C	C	D	D	D
3	B	C	D	C	D	A	A	B	D	A	B	C
4	B	C	D	C	D	A	A	B	D	A	B	C
5	C	D	B	D	A	C	B	D	A	B	C	A
⋮	⋮											

The experimenter was intrigued with design 2 since a subject would remain on each of the diets for a relatively long period (50 days). Note that for three periods both designs 1 and 2 are balanced incomplete block (BIB) in columns, the former being a binary BIB and the latter a ternary BIB. These are Youden rectangle

designs to measure direct and residual effects of the treatments.

Since the experimenter was uncertain about the necessary length of a treatment period to obtain stabilized responses, and since he suspected that the period could be shorter than 25 days, he decided to use design 1 for males 1 to 12 and the following design (design 3) for males 13 to 24 and to obtain as many periods as possible:

	males											
Period	13	14	15	16	17	18	19	20	21	22	23	24
1	B	C	D	A	C	D	A	B	D	A	B	C
2	A	A	A	B	B	B	C	C	C	D	D	D
3	A	A	A	B	B	B	C	C	C	D	D	D
4	A	A	A	B	B	B	C	C	C	D	D	D
5	A	A	A	B	B	B	C	C	C	D	D	D

For three periods design 1 is variance optimal for both residual and direct treatment effects; design 2 is the least variance optimal BIB design possible. Also, design 3 is one of the least variance optimal designs, but it is better than design 2. However, from a nutritional and medical standpoint, design 3 is the optimal BIB design because an individual will remain on a diet as long as is possible in this experiment, and any doubts about diet and response stabilization that a nutritional or medical researcher might have will be satisfied. This illustrates that subject matter and statistical optimality can be at odds. Using both designs 1 and 3 capitalizes on both kinds of optimality. In experimentation, it is often necessary to satisfy non-statistical criteria in order to have the experimental results accepted by others in the field.

During the conduct of the experiment in period 1, it was found that diet and response stabilization was achieved in less than 14 days. Hence, the length of

the treatment period was set at two weeks with daily and weekly measurements being taken on many responses, and with selective samplings on several more responses. In addition, it was found that some males were losing weight. Rack candy and fat cookies (fiber-free) were used to attempt weight stabilization. This was an unsuccessful solution, and during periods 4 and 5 it was decided to increase the diets by 0%, 21%, 42%, or 63%, depending upon the amount needed for weight stabilization. The two designs actually used for weekly measurements were:

			males (design 4)											
Period		Week	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1/10-16	A	B	C	D	A	B	C	D	A	B	C	D
	2	1/17-23	A	B	C	D	A	B	C	D	A	B	C	D
2	1	1/24-30	B	A	D	C	D	C	B	A	C	D	A	B
	2	1/31-2/6	B	A	D	C	D	C	B	A	C	D	A	B
3	1	2/7-13	C	D	A	B	B	A	D	C	D	C	B	A
	2	2/14-20	C	D	A	B	B	A	D	C	D	C	B	A
% increase in diet			42	42	42	63	0	63	42	42	21	42	21	63
4	1	2/21-27	D	C	B	A	C	D	A	B	B	A	D	C
	2	2/28-3/6	D	C	B	A	C	D	A	B	B	A	D	C
5	1	3/7-13	E	E	E	E	E	E	E	E	E	E	E	E
	2	3/14-20	E	E	E	E	E	E	E	E	E	E	E	E
	3	3/21-27	E	E	E	E	E	E	E	E	E	E	E	E

males (design 5)

Period	Week		13	14	15	16	17	18	19	20	21	22	23	24
1	1	1/10-16	B	C	D	A	C	D	A	B	D	A	B	C
	2	1/17-23	B	C	D	A	C	D	A	B	D	A	B	C
2	1	1/24-30	A	A	A	B	B	B	C	C	C	D	D	D
	2	1/31-2/6	A	A	A	B	B	B	C	C	C	D	D	D
3	1	2/7-13	A	A	A	B	B	B	C	C	C	D	D	D
	2	2/14-20	A	A	A	B	B	B	C	C	C	D	D	D
% increase in diet			42	63	42	42	42	42	21	63	21	0	63	21
4	1	2/21-27	A	A	A	B	B	B	C	C	C	D	D	D
	2	2/28-3/6	A	A	A	B	B	B	C	C	C	D	D	D
5*	1	3/7-13	A	A	A	B	B	B	C	C	C	D	D	D
	2	3/14-20	A	A	A	B	B	B	C	C	C	D	D	D
	3	3/21-27	A	A	A	B	B	B	C	C	C	D	D	D

*During period 5 subjects received an additional increment of fiber equal to the levels appearing in periods 1 to 3.

Some characteristics were measured only bi-weekly on the even numbered weeks, others were taken weekly, and others were recorded on the second, the fourth, and weekly thereafter. Also, some measurements were affected by the change in the experiment for periods 4 and 5, and some were not. Thus a variety of statistical analyses will be required for all characteristics observed.

2.4. Response Equation

For many of the measurements collected or for some transformation of these measurements, we shall consider the following response model to be a first approximation to the true model for the designs in Section 2.3:

$$Y_{hijk} = \mu + \rho_h + \gamma_i + \delta_j + \pi_k + \epsilon_{hijk} , \qquad (2.1)$$

where μ is a common effect, ρ_h is the effect of the h^{th} period, γ_i is the effect of the i^{th} individual, δ_j is the direct effect of treatment j, π_k is the residual effect of treatment k in the period following the one in which it was applied, and the ϵ_{hijk} are random independent normal deviates with mean zero and common variance σ^2_ϵ. Superscript h runs from 1 to p where p depends upon how many periods or weeks are included in an analysis; subscript i runs from 1 to 12 for design 4 and from 13 to 24 for design 5. Subscripts j and k take on the values 1 = A, 2 = B, 3 = C, 4 = D, and 5 = E. If boys 1 to 12 were blocked into groups of four by latin squares, analyses could be performed on each group of four males and subscript i would be replaced by gi and the γ_i effect would be replaced by $\alpha_g + \gamma_{gi}$ in (2.1). Analyses may also be made for groups of 12 or 24 males. We shall consider analyses in groups of 12 for various kinds of observations on designs 4 and 5.

2.5. Statistical Analyses

The first four periods of design 4 can be analyzed by standard methods (see Federer [1955], chapter XIV) when the measurements are unaffected by the change in diet amount in period 4.

Given homoscedasticity, normal equations for (2.1) in matrix form are:

$$\left[\begin{array}{cc:cc} 12I_{p\times p} & J_{p\times 12} & & P' \\ J_{12\times p} & PI_{12\times 12} & & C' \\ \hdashline & & r_d I_{v\times v} & T_{v\times v} \\ P & C & T_{v\times v} & r_r I_{v\times v} \end{array}\right] \left[\begin{array}{c} \underline{\rho+\mu} \\ \underline{\gamma} \\ \underline{\delta} \end{array}\right] = \left[\begin{array}{c} \underline{Y}_p \\ \underline{Y}_c \\ \underline{Y}_d \end{array}\right] , \qquad (2.2)$$

where the various matrices J, P, C, and T are incidence matrices between pairs of categories, J is a matrix of ones indicating orthogonality of periods (rows) and subjects (columns) in the design, I is an identity matrix, the prime indicates the transpose of a matrix, $\underline{Y}_p$ is a vector of row totals, $\underline{Y}_c$ is a vector of column totals, $\underline{Y}_d$ is a vector of treatment responses in the period in which it was applied (direct effects), $\underline{Y}_r$ is a vector of totals of responses in the period following the period in which the treatment was applied (residual effects), r_d is the number of replicates for direct effects and r_r is the number for residual effects. Eliminating period (row) and subject (column) effects results in the following reduced normal equations:

$$\left[\begin{bmatrix} r_d I_{v\times v} & T'_{v\times v} \\ T_{v\times v} & r_r I_{v\times v} \end{bmatrix} - \frac{1}{12} pp' - \frac{1}{p} CC'\right] \cdot \begin{bmatrix} \underline{\delta} \\ \underline{\pi} \end{bmatrix}$$

$$= \begin{bmatrix} \underline{Y}_d \\ \underline{Y}_r \end{bmatrix} - \frac{1}{12} P\underline{Y}_p - \frac{1}{p} C\underline{Y}_c = \begin{bmatrix} \underline{Q}_d \\ \underline{Q}_r \end{bmatrix} . \qquad (2.3)$$

Applying the constraints $\Sigma\hat{\rho}_h = \Sigma\hat{\gamma}_i = \Sigma\hat{\delta}_j = \Sigma\hat{\pi}_k = 0$, results in unique solutions for the effects when the designs are connected. Design 4 for 5 periods is not connected since the contrast of E versus rest is also a period contrast.

For the first three periods of design 1, the solution for $\hat{\mu}$ is $\bar{y}$ and for $\hat{\rho}_h$ is $(\bar{y}_{h\ldots} - \bar{y})$; the solutions for $\hat{\delta}_j$ and $\hat{\pi}_k$ are (for I_4 = 4x4 identity matrix) obtained from (2.3). The variance of a difference between any two $\hat{\delta}_j$ is $21\sigma_\epsilon^2/68$ and of a difference between any two $\hat{\pi}_k$ is $9\sigma_\epsilon^2/17$. The nonorthogonality reduced the effective replication from 9 to 6.5 for direct effects and from 6 to 3.8 for residual effects

in design 1 with 3 periods.

For the first 3 periods of design 3, the solutions for the $\underline{\delta}$ and $\underline{\pi}$ are:

$$\begin{bmatrix} \hat{\underline{\delta}} \\ \hat{\underline{\pi}} \end{bmatrix} = \begin{bmatrix} 3I_4/16 & 0 \\ 0 & 3I_4/14 \end{bmatrix} \cdot \begin{bmatrix} \underline{Q}_d \\ \underline{Q}_r \end{bmatrix} \tag{2.4}$$

and for all 5 periods the solutions are:

$$\begin{bmatrix} \hat{\underline{\delta}} \\ \hat{\underline{\pi}} \end{bmatrix} = \frac{5}{272} \begin{bmatrix} 9I_4 & -2I_4 \\ -2I_4 & 8I_4 \end{bmatrix} \cdot \begin{bmatrix} \underline{Q}_d \\ \underline{Q}_r \end{bmatrix} . \tag{2.5}$$

For design 3 with 3 periods the effective replication has been reduced from 9 to 5.3 for direct effects and from 6 to 4.7 for residual effects; for 5 periods, the effective replication was reduced from 15 to 6.0 for direct effects and from 12 to 6.8 for residual effects. The additional 6 replications only increase the effective replication by 0.7 for direct effects and by 2.1 for residual effects.

For design 4 with weekly measurements, i.e., 6 periods, the solutions for the $\underline{\delta}$ and $\underline{\pi}$ are:

$$\begin{bmatrix} \hat{\underline{\delta}} \\ \hat{\underline{\pi}} \end{bmatrix} = \frac{3}{3280} \begin{bmatrix} 79I_4 & -32I_4 \\ -32I_4 & 96I_4 \end{bmatrix} \cdot \begin{bmatrix} \underline{Q}_d \\ \underline{Q}_r \end{bmatrix} . \tag{2.6}$$

For design 5, the solutions for the $\underline{\delta}$ and $\underline{\pi}$ for 6 periods and for 11 periods are, respectively:

$$\begin{bmatrix} \hat{\underline{\delta}} \\ \hat{\underline{\pi}} \end{bmatrix} = \frac{3}{1504} \begin{bmatrix} 63I_4 & -32I_4 \\ -32I_4 & 64I_4 \end{bmatrix} \cdot \begin{bmatrix} \underline{Q}_d \\ \underline{Q}_r \end{bmatrix} \qquad (2.7)$$

and

$$\begin{bmatrix} \hat{\underline{\delta}} \\ \hat{\underline{\pi}} \end{bmatrix} = \frac{11}{7144} \begin{bmatrix} 79I_4 & -46I_4 \\ -46I_4 & 72I_4 \end{bmatrix} \cdot \begin{bmatrix} \underline{Q}_d \\ \underline{Q}_r \end{bmatrix} \qquad (2.8)$$

Similarly, one may obtain solutions for various other types of situations.

It should be noted that the change in the amount of food given each boy starting with week 7, induces a column x row (subject x period) interaction for some characteristics. Thus, an additional set of 8 normal equations will need to be included in (2.2) as obtained from the following table:

	Design 4				Design 5			
	boys receiving diet increments of				boys receiving diet increments of			
	0%	21%	42%	63%	0%	21%	42%	63%
Periods 1,2,3	No. 5	Nos. 9,11	Nos. 1-3, 7,8,10	Nos. 4, 6,12	No. 22	Nos. 19, 21,24	Nos. 13, 15-18	Nos. 14, 22,23
Period 4								

An analysis of variance for design 5 would be:

Sources of variation	d.f.	
Correction for mean	1	
Subjects	11	
Periods	4	
Subjects x periods	44	
Subject x period (ign. tr.)	3	
Remainder	41	
Direct (elim. Subjects and period x subject; ign. res.)		3
Residual (elim. all else)		3
Remainder		35

Data for percent digestibility of cell wall fiber was obtained. The arcsine transformations of these percentages are given in Table 2.1. For design 4 (boys 1-12), data were obtained for weeks 2, 4, 6, 8 for treatments A, B, C, and D. For design 5 (boys 13-24), data were obtained for weeks 2, 4, 5, 6, 7, 8, 9, 10, 11. For these data, it was necessary to modify equation (2.1) since there was a continuing effect for treatment C. This treatment tended to eliminate intestinal bacteria so that digestibility percentage was lowered and evidently the bacteria level was not restored (see, e.g., data for boy 10). With this additional component in (2.1), analyses of variance are presented for design 4, weeks 2, 4, 6, and for design 5, weeks 2, 4, 5, 6, and weeks 2, 4-11 in Tables 2.2 to 2.4. It should be noted that in design 4, weeks 2, 4, 6, the carry-over effect is for a two-week period, whereas that for design 5 is for a one-week period.

Table 2.1. Arcsine of $\sqrt{\text{digestibility percent of cell wall fiber}}$ for 24 subjects over a period of 11 weeks.

		Subject (design 4)												
Period	Week No.	1	2	3	4	5	6	7	8	9	10	11	12	Sum
1	2	53A	39B	27C	56D	47A	48B	38C	71D	45A	48B	24C	57D	555
1	4	54B	46A	62D	27C	64D	35C	46B	46A	29C	64D	51A	46B	570
3	6	26C	54D	45A	48B	49B	43A	66D	31C	29D	10C	46B	42A	489
5	10	63E	61E	68E	59E	48E	63E	70E	59E	67E	31E	63E	59E	711
	11	70E	64E	67E	70E	57E	67E	66E	66E	61E	18E	69E	54E	729
Sum (wks 2,4,6)		133	139	136	131	160	126	150	148	103	122	121	145	1614
wks 8,10		130	89	117	107	81	128	126	107	109	66	128	99	1287
wks 2,4,6,8,10		263	228	253	238	241	254	276	255	212	188	249	244	2901
Sum (all wks)		333	292	320	308	298	321	342	321	273	206	318	298	3630

		Subject (design 5)												
Period	Week No.	13	14	15	16	17	18	19	20	21	22	23	24	Sum
1	2	46B	19C	75D	39A	19C	71D	35A	48B	57D	46A	47B	29C	531
2	4	47A	37A	38A	37B	51B	47B	10C	54C	17C	65D	57D	67D	527
3	5	44A	36A	51A	42B	40B	46B	22C	23C	27C	66D	67D	69D	533
	6	46A	41A	46A	36B	46B	46B	16C	39C	13C	73D	63D	64D	529
4	7	47A	43A	43A	35B	46B	54B	17C	32C	29C	69D	68D	61D	544
	8	47A	42A	44A	29B	48B	48B	16C	29C	16C	66D	62D	66D	513
5	9	44A	38A	46A	39B	47B	47B	19C	15C	20C	67D	65D	71D	518
	10	46A	40A	47A	42B	45B	46B	20C	14C	19C	69D	66D	66D	520
	11	46A	39A	46A	41B	47B	46B	20C	13C	19C	72D	70D	66D	525
Sum (wks 2,4,6)		139	97	159	112	116	164	61	141	87	184	167	160	1587
Sum (all wks)		413	335	436	340	389	451	175	267	217	593	565	559	4740
wk 2,4,6,8,10		232	179	250	183	209	258	97	184	122	319	295	292	2620

3. EXAMPLES OF VARIATIONS OF REPEATED MEASURES DESIGNS

Five examples of variations of repeated measurements designs are presented below. It should be noted that each situation demands a different variation. This illustrates the necessity of designing for the experiment rather than making the experiment fit a known design. Flexibility in design and in analyses is required to meet the needs of the experiment.

Example 3.1. Feinberg, et al. [1976] discuss random and balanced repeated measures designs and two different models. They discuss these designs in the context of so-called social experiments and relate them to a particular study called the Kansas City Preventive Patrol Experiment (KCPPE). The treatments were A, α = search strategy, B, β = visibility strategy, and C, γ = reactive no-patrol strategy and were designed to attempt to reduce crime. The latin letters were used in the primary areas (beats) and the greek letter strategies would be used in secondary areas. The primary area would be surrounded by secondary areas. The primary plus the secondary beats formed a block. There were 3 one-year periods involved. The two designs they discuss are given in Table 3.1. Analyses of variance are also presented. Their model 1 corresponds to (2.1) for each set of letters, i.e., A, B, C, and α, β, γ. In addition, they obtain an interaction between the two sets. Their model 2 requires all possible pairs and reciprocals, i.e., AB vs BA, AC vs CA, BC vs CB to obtain the 3 degrees of freedom for "Primary 'carry-over'" effects. The second model does not require the carry-over effect of C to A to be the same as C to B, whereas equation (2.1) does.

Table 3.1. Designs and analyses of variance for KCPPE

Block	Random design Time period T_1	T_2	T_3	Balanced design Time period T_1	T_2	T_3
1	A γ	B β	C α	A α	B β	C γ
2	C γ	A β	B α	B γ	C α	A β
3	B β	A γ	C α	C β	A γ	B α
4	C β	B α	A γ	B α	A γ	C β
5	B α	C γ	A β	C γ	B β	A α
6	B β	C α	A γ	A β	C α	B γ
7	C β	A α	B γ	A γ	B β	C α
8	B α	A β	C γ	B α	C γ	A β
9	B β	C γ	A α	C β	A α	B γ
10	C α	A γ	B β	A α	C β	B γ
11	A α	B β	C γ	B β	A γ	C α
12	C γ	B α	A β	C γ	B α	A β

Source of variation	d.f.	
Correction for mean	1	
Blocks	11	
Time (periods)	2	
Primary strategy	2	
Secondary strategy	2	
Primary × secondary	4	
Primary residual	2	their model 1
Secondary residual	2	their model 1
Remainder	10	their model 1
Primary "carry-over"	3	their model 2
Secondary "carry-over"	3	their model 2
Remainder	8	their model 2

Example 3.2. Four treatments, A, B, C, and D, were selected for the experiment, and formed a 2 x 2 factorial treatment design as follows:

Level of Factor One (a)	Level of Factor Two (b)	
	b_0	b_1
a_0	A	B
a_1	C	D

Since it was planned to apply the four treatments sequentially to each sampling unit, it was considered likely that there might be a one-period carry-over effect.

The experiment design selected was:

	Male Students											
	square 1				square 2				square 3			
Order	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	B	A	D	C	B	A	D	C
3	C	D	A	B	C	D	A	B	C	D	A	B
4	D	C	B	A	D	C	B	A	D	C	B	A

	Female Students											
	square 4				square 5				square 6			
Order	13	14	15	16	17	18	19	20	21	22	23	24
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	B	A	D	C	B	A	D	C
3	C	D	A	B	C	D	A	B	C	D	A	B
4	D	C	B	A	D	C	B	A	D	C	B	A

Each square forms the same latin square design of order four. This square allows solutions for residual effects of the treatments to be obtained.

Using response equation (2.1), one obtains analyses of variance for each of the 6 latin squares as given in Table 3.2. Summing over the various lines in Table 3.2, one can use the sums to form parts of Tables 3.3 and 3.4 as indicated. These sums are then partitioned into the sources of variation indicated.

Table 3.2. Analyses of variance for 6 latin squares.

Source of Variation	Degrees of Freedom						
	Square 1	Square 2	Square 3	Square 4	Square 5	Square 6	Total
Total	16	16	16	16	16	16	96
Correction for mean	1	1	1	1	1	1	6
S = Sequence (column)	3	3	3	3	3	3	18
D = Direct effect of treatment (ign. res. eff.)	3	3	3	3	3	3	18
S x D	9	9	9	9	9	9	54
Orders (ign. res. eff.)	3	3	3	3	3	3	18
Residual effect (elim. all other effects)	3	3	3	3	3	3	18
Remainder	3	3	3	3	3	3	18
Direct effects (elim. res. eff.)	3	3	3	3	3	3	18

Example 3.3. Drug X was a standard drug and drug Y was a new drug whose performance was to be compared with drug X. It was decided to use four treatment periods on each individual and to use the two sequences XYXY and YXYX in the four periods. The experiment design was a simple change-over design with four periods. For medical and practical reasons, it was decided to use n_1 individuals for sequence XYXY and n_2 individuals for the second sequence of treatments, i.e., YXYX.

Using response equation (2.1), Table 3.5 may be obtained. An alternate form of an analysis of variance is given in Table 3.6.

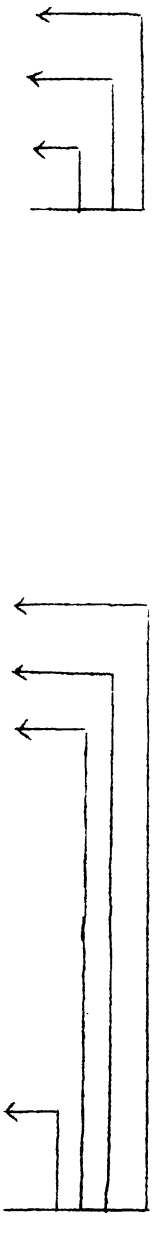
Table 3.3. A pooled ANOVA for the six separate analyses for each square.

Source of Variation	Degrees of Freedom	Sum of Items in Table 3.2 for	F-tests (see arrows)
Total	96	line 1	
Correction for mean	1		
Sex	1	line 2	
Squares within sex	4		
Sequences within squares	18	line 3	
Sequence	3		
Sequence × sex	3		
Sequence × square wn sex	12	pool for error*	
Direct (ignoring residual) wn sq.	18	line 4	
Direct (ignoring residual)	3	not a useful partition if residual effects are present *	
Direct (ign. res.) × sex	3		
Direct (ign. res.) × sq. wn sex	12		
Sex × direct within squares	54	line 5	
Orders (ign. res.) within squares	18	line 6 (same as *)	
Residual (elim. dir. and order) wn sq.	18	line 7	
Residual (elim. dir. and order)	3		
Residual × sex	3		
Residual × squares within sex	12	error	
Remainder within squares	18	line 8	
Direct (elim. res.) within squares	18		
Direct (eliminating residual)	3		
Direct (elim. res.) × sex	3		
Direct (elim. res.) × squares wn sex	12		

* Only appropriate if there are no residual effects.

Table 3.4. An alternate pooled ANOVA for the six squares.

Source of Variation	Degrees of Freedom			Sum of Items In	F-tests (see arrows)
Total	96			line 1	
Correction for mean	1			sum of lines 2 and 3	
Among students	23				
Sex		1			
Sequence		3			
Sex × sequence		3			
Students within sex and square		16 *			
Males			8		
Females			8		
Within students	72				
Direct (ignoring residual) wn sq.		18		line 4	
Direct (ign. res.) × sequence		9			
Order (ignoring residual)			3		
Residual (elim. dir. and order)			3		
Remainder			3		
Direct (ign. res.) × sequence × square		45			
Order (ign. res.) × square			15		
Order (ign. res.) × sex			3		
Order (ign. res.) × square wn sex			12		
Residual (elim. dir. and order) × square			15		
Residual (elim. all else) × sex			3		
Residual (elim. all else) × sq. wn sex			12	error	
Remainder			15		



* Only appropriate if there are no residual effects.

Table 3.5. An analysis of variance for a four period change-over design with two treatments.

	Degrees of Freedom	F-tests (see arrows)
Total	$4(n_1+n_2)$	
Correction for mean (ign. all else)	1	
Period (elim. mean; ign. all else)	3	
Patient (elim. mean; ign. direct and residual)	n_1+n_2-1	
Direct (elim. mean, period, patient; ign. residual)	1	
Residual (elim. all else)	1	←
Remainder	$3(n_1+n_2)-5$	—
Direct (elim. all else)	1	←

Table 3.6. An alternative ANOVA when no residual effects are present.

Source of Variation	Degrees of Freedom	F-tests (see arrows)
Total	$4(n_1+n_2)$	
Correction for mean	1	
Patients = P	n_1+n_2-1	
Sequence	1	←
Within sequence	n_1+n_2-2	
Within XYXY	n_1-1	
Within YXYX	n_2-1	
Period	3	
Linear = L	1	←
Quadratic = Q	1	←
Cubic = C	1	←
P × period	$3(n_1+n_2-1)$	
P × L	n_1+n_2-1	
L × sequence	1	←
Remainder	n_1+n_2-2	
P × Q	n_1+n_2-1	
Q × sequence	1	←
Remainder	n_1+n_2-2	
P × C	n_1+n_2-1	
C × sequence	1	←
Remainder	n_1+n_2-2	

Example 3.4. In some experimental investigations it becomes necessary to alter the randomization requirements of the experiment design. Such was the case in a study of asthma attacks in humans wherein two drugs and a combination of the two drugs were given to patients to alleviate the effects of an asthma attack. The combination was only given to a patient after the response to the individual drugs was known.

The treatment design consisted of the following four treatments for each of two mediators, histamine and methacholine:

Placebo — Denoted by the symbol P.
Scholl's 1000-BR — Denoted by the symbol S.
Isoproterenol — Denoted by the symbol I.
Combination = S + I — Denoted by the symbol C.

A mediator is a chemical used to induce an asthma attack. Since asthma attacks can be fatal, it is necessary to monitor the procedure very carefully. The treatment design could be considered as a 2^3 factorial of the following nature:

Histamine

Level of I	Level of S 0	1
0		
1		

Methacholine

Level of I	Level of S 0	1
0		
1		

where 0 means the drug was absent and 1 means the presence of the drug in the treatment. This treatment design represents the simplest possible design for mixtures of treatments.

The experimental design (plan) used is given in Table 3.7. Treatment P was given on visits one and two; the doctor administering the mediator and treatment knew the identity of the treatment on these two visits but not on visits three through six; and the patient did not know the identity of the treatment or mediator being administered on all six visits. Thus, the study was singly-blind on visits one and two and double-blind on visits three through six. The four treatments were randomly allotted on visits three through six, except C could not be given until both S and I had been given. This means that C could not appear on visits three and four and that S and I cannot appear on visit six. In addition, after the first half of the patients had received the treatments as described above, it was decided to use additional sequences as new patients were added to the study. Thus, eight sequences of treatments resulted and each sequence appeared twice except for s_4 which appeared on three patients. Note that for the doubly-blind part of the experiment, visits three through six, treatments (P,S,I,C), visits (3,4,5,6), and mediators (M,H) are orthogonal to patients and that visits and treatments are nonorthogonal.

An analysis of variance is presented in Table 3.8 for the data from all four visits. Analyses of variance tables for data from visits 3 and 5 (histamine mediator) and from visits 4 and 6 (methacholine mediator) are given in Table 3.9. Then, one form of a pooled analysis of variance is given in Table 3.10.

Table 3.7. Experiment design for administration of drugs and mediators for six visits of each of 17 patients.

Patient	Treatment for visit (no.) and mediator*						Sequence of Treatments
	1 - H	2 - M	3 - H	4 - M	5 - H	6 - M	
1	P	P	S	I	P	C	S_1
2	P	P	S	I	C	P	S_2
3	P	P	I	S	P	C	S_3
4	P	P	P	S	I	C	S_4
5	P	P	I	P	S	C	S_5
6	P	P	I	S	C	P	S_6
7	P	P	S	P	I	C	S_7
8	P	P	P	I	S	C	S_8
9	P	P	I	S	P	C	S_3
10	P	P	P	S	I	C	S_4
11	P	P	I	S	C	P	S_6
12	P	P	S	I	C	P	S_2
13	P	P	S	I	P	C	S_1
14	P	P	P	I	S	C	S_8
15	P	P	S	P	I	C	S_7
16	P	P	I	P	S	C	S_5
17	P	P	P	S	I	C	S_4

* H denotes use of histamine mediator and M denotes use of methacholine mediator.

Table 3.8. An analysis of variance for the 17 patients given four treatments on four different visits.

Source of Variation	Degrees of Freedom		F-test (see arrows)
Total	68		
Correction for mean	1		
Patients	16		
Sequences		7	←
Within Sequences		9	←
Visits (ignoring treatment)	3		
Histamine vs. Methacholine		1	
Remainder		2	
Treatments (eliminating visits)	3		←
Remainder	45		←
Visit (eliminating treatment)	3		←
3 + 5 vs. 4 + 6 = Histamine vs. methacholine		1	←
Remainder		2	←

Table 3.9. An analysis of variance for data for each mediator for 17 patients and two visits.

Source of Variation	Degrees of Freedom		F-test (see arrows)
	Visits 3 and 5	Visits 4 and 6	
Total	34	34	
Correction for mean	1	1	
Patients (ign. treatment)	16	16	
Visit (ignoring treatment)	1	1	
Visit x patient	16	16	
Treatment (elim. all else)	3	3	←
Remainder	13	13	←
Visit (elim. treatment)	1	1	←

Table 3.10. A pooled ANOVA using the ANOVAs in Table 3.9.

Source of Variation	Degrees of Freedom		F-test (see arrows)
Total	68		
Correction for mean	1		
Patients	16		
Sequences		7	←
Within sequence		9	←
Visits (ignoring treatment)	3		
Treatments (elim. visits and ign. interactions)	3		
Treatment × mediator plus patient × mediator (elim. visit, treatment, and patient)	19		←
Remainder	26		←
Within histamine		13	
Within methacholine		13	
Treatment (elim. all else)	3		←
Visit (elim. all else)	3		←

Example 3.5. A psychology graduate student was interested in the effect of choice on performance for various types of tests and for boys and girls in an elementary school. The nature of the administration of the tests dictated the type of experimental design that could be used. Three types (A = English, B = mathematics, and C = music) of tests with two parts (general = G and technical = T) for each type were to be used in the investigation. In

addition, a third factor, choice (c) versus no-choice (n) on which part (G or T) the student takes, was to be used. The fourth factor in the experiment was to be sex as girls may respond differently from boys. It is believed that there may be a residual effect of the type of test and interaction for direct effects of type of tests and parts of the test, and that pupils would score higher when given a choice of which part of the test they would take.

The investigator could administer the type and part of a test to only one student at a time, and since she wanted all students to take all tests, she used the following experiment design for 12 girl students.

	Girl Pupil (Number)											
Order	1	2	3	4	5	6	7	8	9	10	11	12
1	cAX	nAX	cBX	nBX	cCX	nCX	cAS	nAX	cBX	nBX	cCX	nCX
2	nB?	cB?	nC?	cC?	nA?	cA?	nC?	cC?	nA?	cA?	nB?	cB?
3	cCX	nCX	cAX	nAX	cBX	nBX	cBX	nBX	cCX	nCX	cAX	nAX

Pupils 1 and 2 have the same sequence of type of test (i.e., ABC), have opposite sequences of choices (i.e., cnc versus ncn), have the same parts X of the test in periods one and three (i.e., whatever part student one chooses in periods one and three, student two must have these same parts in these two periods), and may have the same or different parts of a test in period two, depending upon the choice of student two. Similar arrangements exist between members of the other five pairs of pupils (i.e., 3 and 4, 5 and 6, 7 and 8, 9 and 10, and 11 and 12). This procedure was used to obtain as precise a contrast as possible on choice versus no-choice. The numbers 1,2,···,12 were randomly assigned to the 12 girl pupils. A similar plan was used for the 12 boy pupils. It was considered

impractical to use more than three or four periods because of fatigue.

It was believed that the response, or some function of the response, could be properly formulated as a linear model of the form:

$$Y_{efghijk} = \mu + \rho_e + \lambda_f + \gamma_{gf} + \alpha_h + \beta_i + \delta_j + \pi_k + \beta\delta_{ij} + \epsilon_{efghijk} ,$$

where μ is an effect common to all observations, ρ_e is the effect of the e^{th} ($e = 1,2,3$) order of administering a test, λ_f is the effect of the f^{th} sex ($f = \sigma^{\prime}, \varphi$), γ_{gf} is the effect of the g^{th} ($g = 1,2,\cdots,12$) student plus sequence effect, α_h is the effect of the h^{th} ($h = c,n$) choice, β_i is the direct effect of the i^{th} ($i = A,B,C$) type of test, δ_j is the effect of the j^{th} ($j = G,T$) part of a test, π_k is the one-period residual effect of the k^{th} ($k = A,B,C$) type of test, $\beta\delta_{ij}$ is an interaction effect of the part of a test with the direct effect of type of test, and the $\epsilon_{efghijk}$ are $NIID(0,\sigma_\epsilon^2)$.

An analysis of variance and various F-tests could be obtained for data from either the 12 girls or for the 12 boys (Table 3.11). The data are highly non-orthogonal and a series of eliminations from the set of 32 normal equations would be needed. A much simpler procedure for testing the responses to choice versus no-choice exists, if one is willing to consider only a part of the data. If one considers the pairs of pupils and observes the facts that the differences $Y_{1fgcijk} - Y_{1fg'nijk} = \alpha_c - \alpha_n + \gamma_g - \gamma_{g'} + \epsilon_{efgcijk} - \epsilon_{efg'nijk} = d_{1m}$ and that $Y_{3fgcijk} - Y_{3fg'nijk} = \alpha_c - \alpha_n + \gamma_g - \gamma_{g'} + \epsilon_{efgcijk} - \epsilon_{efg'nijk} = d_{3m}$ for each pair g and g' of pupils. Then, we may construct the following table of differences:

Order	Pair 1	2	3	4	5	6	Total
1	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	$d_{1.}$
3	d_{31}	d_{32}	d_{33}	d_{34}	d_{35}	d_{36}	$d_{3.}$
							$d_{..}$

An analysis of variance on these 12 differences would provide a test for the contrast of choice versus no-choice as follows:

Source of Variation	d.f.	Sum of Squares	F-test (see arrows)
Total	12	$\sum_{m} \sum_{e=1,3} d^2_{em}$	
Choice vs. no-choice	1	$d^2_{..}/12$	←
Order × choice	1	$(d_{1.}-d_{3.})^2/12$	←
Within order	10	$\sum_{e} \left(\sum_{m=1} d^2_{em} - d^2_{e.}/6 \right)$	
Pair differences	5		
Remainder	5		

This test would not have the power of the one given in Table 3.11, but simplicity of computation and the presence of a relatively large effect for choice versus no-choice would increase the desirability of using the procedure. Similar tests can be made for other effects by pairing observations similar in all respects but the one under consideration and student. It should be noted that student to

Table 3.11. An ANOVA for girl pupils (or for boy pupils).

Source of Variation	Degrees of Freedom	F-test (see arrows)
Total	36	
Correction for mean (ignoring all below)	1	
Orders (elim. above; ignoring effects below)	2	
Students (elim. above; ignoring effects below)	11	
Types (elim. above; ignoring effects below)	2	
Parts (elim. above; ignoring effects below)	1	
Types x parts (elim. above; ign. effects below)	2	
Residual for types (elim. all above; ign. choice)	2	
Choice (eliminating all above)	1	←
Remainder	14	←
Residual for types (elim. all else)	2	←
Types x parts (elim. all else)	2	←
Parts (elim. all else)	1	←
Types (elim. all else)	2	←

student differences do not appear in the order × pair interaction of the above d_{em}.

The test for sex effect can be made using the above procedure and pairing a boy and a girl treated similarly in all respects. Then, for all the p resulting pairs, a simple t-test with p-1 degrees of freedom, may be utilized. One unsatisfactory procedure that comes to mind is to combine all data from both boys and girls and from the resulting 34 normal equations to obtain solutions of effects and to compute the following sums of squares:

Sex (eliminating all other effects except pupil wn sex)

Pupil (eliminating all other effects) .

Then, from the resulting mean squares, compute an F-statistic of the ratio of the former to the latter mean square. This is unsatisfactory because the mean squares are correlated. A second procedure which has the same difficulty as the above one is to correct each student total for the effects in the total, construct a one-way ANOVA for between sex and for pupils within sex, and then compute an F-statistic. However, as before, the resulting mean squares are correlated.

4. DISCUSSION AND CONCLUSIONS

Several variations of two-way, repeated measurements designs for estimating residual and direct treatment effects have been considered. A study of these examples leads us to the following conclusions:

(i) Statistical textbook, and perhaps statistical literature, coverage is inadequate for many real world situations.

(ii) The sampling unit, the sampling procedure, and the experimental unit must be precisely defined and understood.

(iii) It is essential for both the experimenter and the statistician to fully comprehend the sampling procedure and the experiment design used in the investigation.

(iv) Experimental conditions often dictate the statistical design.

(v) The response model may be unknown and a linear model representation can only be considered as an approximation to the true situation.

(vi) It is essential that both the experimenter and the statistician fully comprehend the nature of the response model used and the statistical analyses made.

(vii) Many experimenters and statisticians may be unaware of alternate models and analyses in published literature.

(viii) The length of a period may determine whether or not residual effects are present.

(ix) The types for residual effects may differ from expectation.

A technique that has been found to be useful over the years in connection with items (iii), (v), and (vi) above, is to construct an artificial example both with and without effor components. To illustrate, consider design 1 in section 2. Let the values for the parameters of the yield equation (2.1) be:

$$\mu = 10,\ \alpha_1 = \alpha_2 = \alpha_3 = 0,\ \rho_1 = -2,\ \rho_3 = 0,$$

$$\gamma_{11} = -1 = \gamma_{12},\ \gamma_{13} = 1 = \gamma_{14},\ \text{all other } \gamma_{gi} = 0,$$

$$\delta_A = -3,\ \delta_B = -2,\ \delta_C = 1,\ \delta_D = 4,\ \pi_A = -1 = \pi_B,\ \pi_C = 2,\ \pi_D = 0,$$

$$\text{all } \epsilon_{ghijk} = 0.$$

Then, the values for the Y_{ghijk} are:

Period	Student 1	2	3	4	5	6	7	8	9	10	11	12	Total	Mean
1	4	5	10	13	5	6	9	12	5	6	9	12	96	$8 = \bar{y}_{.1...}$
2	8	7	19	14	15	12	12	9	12	15	11	10	144	$12 = \bar{y}_{.2...}$
3	9	12	8	11	8	9	13	10	16	11	7	6	120	$10 = \bar{y}_{.3...}$
Total	21	24	37	38	28	27	34	31	33	32	27	28	360 $= Y_{.....}$	$10 = \bar{y}$
Total for square				120				120				120		

$Y_{...A.} = 64$ $Y_{...C.} = 96$ $Y_{....A} = 64$ $Y_{....C} = 78$

$Y_{...B.} = 73$ $Y_{...D.} = 127$ $Y_{....B} = 62$ $Y_{....D} = 60$

For design two, let the values for the parameters of the yield equation (2.1) be:

$$\mu = 10,\ \rho_1 = -2,\ \rho_2 = 2,\ \rho_3 = 0,$$

$$\gamma_{13} = -1 = \gamma_{14},\ \gamma_{15} = 1 = \gamma_{16},\ \text{all other } \gamma_i = 0,$$

$$\delta_A = -3,\ \delta_B = -2,\ \delta_C = 1,\ \delta_D = 4,\ \pi_A = -1 = \pi_B,\ \pi_C = 2,\ \pi_D = 0,$$

$$\epsilon_{113A} = 1,\ \epsilon_{213AA} = -1,\ \epsilon_{114A} = -1,\ \epsilon_{214AA} = 1,\ \text{all other } \epsilon_{hijk} = 0.$$

Then the values for the Y_{hijk} are:

	Student													
Period	13	14	15	16	17	18	19	20	21	22	23	24	Total	Mean
1	5	3	6	7	6	6	9	9	9	12	12	12	96	$8 = \bar{y}_{1...}$
2	6	8	9	10	9	9	15	15	15	16	16	16	144	$12 = \bar{y}_{2...}$
3	6	9	14	11	13	6	9	10	16	7	8	11	120	$10 = \bar{y}_{3...}$
Total	17	20	29	28	28	21	33	34	40	35	36	39	360	10

$Y_{..A.} = 59$ $Y_{..B.} = 103$ $Y_{...A} = 52$ $Y_{...C} = 80$

$Y_{..B.} = 71$ $Y_{..C.} = 127$ $Y_{...B} = 58$ $Y_{...D} = 74$

At every step in the computations, a check is possible. For example, $\bar{y}_{1...}$ above must be equal to 10 - 2 = 8 which it does. Likewise, the overall mean $\mu = \bar{y}$ = 360/36 = 10, $Y_{..A.} = 9(\mu + \delta_A) + 2\gamma_{13} + 2\gamma_{14} + 2\gamma_{15} + \gamma_{18} + \gamma_{19} + \gamma_{22} + 2\pi_A$ = 9(10 - 3) + 2(-1) + 2(-1) + 2(1) + 0 + 0 + 0 _ 2(-1) = 59 as it should, etc. The sums of squares for squares and for remainders in design one should be zero, whereas the remainder sum of squares for design two must be equal to $(-1)^2 + 1^2 + (-1)^2 + 1^2 + 0 = 4$. Also, for <u>any</u> computer program used, it is suggested that an example such as the above be used to check the results. If this procedure were followed by all investigators, the quality of published literature would be greatly increased.

Turning our attention now to (v) and (vii), it has been noted that some statisticians have recommended that designs such as the above be discontinued and that the sampling unit be the experimental unit. They do this because they

have noted that students tire at different rates, that milk cows are in different stages of lactation and/or have different lactation curves, that machines wear out at different rates, etc. They apparently are unaware that statistical analyses for latin square designs with differential gradients in the columns have been developed by Cox [1958]. If the above conditions prevail, standard textbook ANOVAs are incorrect, and the statistical analyses given by Cox [1958] should be used. A similar situation exists for split plot designs (see, e.g., Federer [1975]).

Another form of statistical analyses for checking the appropriateness of a given statistical model is to use tests for nonadditivity such as developed, for example, by D. S. Robson, Cornell University, and J. W. Tukey, Princeton University. Suppose that the nonadditivity is of the following form:

Randomized Complete Block Design and Other Two-Dimensional Designs

$$EY_{hi} = \mu + \rho_h + \tau_i \qquad \text{additive model}$$

$$EY^*_{hi} = \mu_{h\cdot}\mu_{\cdot i}/\mu \qquad \text{nonadditive model}$$

$$= \mu + (\mu_{h\cdot}-\mu) + (\mu_{\cdot i}-\mu) + (\mu_{h\cdot}-\mu)(\mu_{\cdot i}-\mu)/\mu$$

$$= \mu + \rho_h + \tau_i + \rho_h\tau_i/\mu \ .$$

In both cases, the tests have been developed considering that the error component is additive, that is, $Y_{hi} - EY_{hi} = \epsilon_{hi}$ and $Y^*_{hi} - EY^*_{hi} = \epsilon^*_{hi}$. Note that if the covariance of the ρ_h and τ_i is zero, both models are identical.

Latin Square Design and Other Three-Dimensional Designs

$$EY_{hij} = \mu + \rho_h + \gamma_i + \delta_j \qquad \text{additive model}$$

$$EY^*_{hij} = \mu_{h\cdot\cdot}\mu_{\cdot i\cdot}\mu_{\cdot\cdot j}/\mu^2 \qquad \text{nonadditive alternate model}$$

$$= \mu + \rho_h + \gamma_i + \delta_j + \frac{1}{\mu}(\rho_h\gamma_i+\rho_h\delta_j+\gamma_i\delta_j+\rho_h\gamma_i\delta_j/\mu) \ ,$$

where

$$\rho_h = \mu_{h\cdot\cdot} - \mu, \ \gamma_i = \mu_{\cdot i\cdot} - \mu, \text{ and } \delta_j = \mu_{\cdot\cdot j} - \mu \ .$$

In both cases the error components are assumed to be additive, i.e., $Y_{hij} - EY_{hij} = \epsilon_{hij}$ and $Y^*_{hij} - EY^*_{hij} = \epsilon^*_{hij}$. In the Tukey test for nonadditivity in a latin square, the term $\rho_h \gamma_i \delta_j / \mu^2$ is omitted. If μ were relatively large, this term would be small relative to the other three terms $\rho_h \gamma_i / \mu$, $\rho_h \delta_j / \mu$, and $\gamma_i \delta_j / \mu$.

Four-Dimensional Designs

$$EY_{hijk} = \mu + \rho_h + \gamma_i + \rho_j + \pi_k \qquad \text{additive model}$$

$$EY^*_{hijk} = \mu_{h\ldots}\mu_{.i.}\mu_{..j}.\mu_{\ldots k}/\mu^3 \qquad \text{nonadditive model}$$

$$= \mu + \rho_h + \gamma_i + \delta_j + \pi_k + (\rho_h\gamma_i + \rho_h\delta_j + \rho_h\pi_k + \gamma_i\delta_j + \gamma_i\pi_k + \delta_j\pi_k)/\mu$$

$$+ (\rho_h\gamma_i\delta_j + \rho_h\gamma_i\pi_k + \rho_h\delta_j\pi_k)/\mu^2 + \rho_h\gamma_i\delta_j\pi_k/\mu^3 ,$$

where $\rho_h = \mu_{h\ldots} - \mu$, $\gamma_i = \mu_{.i.} - \mu$, $\delta_j = \mu_{..j}. - \mu$, and $\pi_k = \mu_{\ldots k} - \mu$. In both cases, the error components might be additive. If so, one could develop various nonadditivity tests depending upon how many terms are deleted from the model. One simply computes the estimated residuals $\hat{\epsilon}_{hijk}$ and $\hat{\epsilon}^*_{hijk}$ and then computes the following one degree-of-freedom statistic:

For each of the above examples, one could compute a test for nonadditivity.

With regard to (viii), the length of a treatment period may determine whether or not a residual effect is present. For example, in marketing experiments on apples, when one day was the length of the period, no carry-over effect existed. When the treatment period was one week, residual effects existed. Apple purchases tend to be once a week purchases. People buying apples on Monday would

not purchase apples on the other days of the week. However, people purchasing too many apples this week would not buy apples next week. Likewise, with regard to the example in section 2, the weekly residual effects are different from the two-week residual effects for a number of characters.

With respect to (ix), the continuing effect of treatment c was unexpected for the example in section 2. The absence of one period carry-over effects was also unexpected. Thus, the statistical analyst must be aware of possible alternative models.

REFERENCES

[1] Atkinson, G. F. "Designs for sequences of treatments with carry-over effects", Biometrics, 22 (1966), 292-309.

[2] Cochran, W. G. and Cox, G. M. Experimental Designs, 2nd ed., New York: John Wiley and Sons, Inc., 1957.

[3] Cox, C. P. "The analysis of Latin square designs with individual curvatures in one direction", Journal of the Royal Statistical Society, Series B, 20 (1958), 193-204.

[4] Cox, D. R. Planning of Experiments, New York: John Wiley and Sons, Inc., 1958, Ch. 13.

[5] Federer, W. T. Experimental Design - Theory and Application, New York: The Macmillan Company, 1955. (Republished by the Oxford and IBH Publishing Co., Calcutta, Bombay, New Delhi, 1967.)

[6] Federer, W. T. "The misunderstood split plot", Proceedings of the Conference at Dalhousie University, Halifax, May 1974, in R. P. Gupta, ed., Applied Statistics, Amsterdam, Oxford: North-Holland Publishing Co., 1975, 9-39.

[7] Federer, W. T. and Balaam, L. N. Bibliography on Experiment and Treatment Design Pre-1968, Edinburgh: Oliver and Boyd, 1972.

[8] Feinberg, S. E., Larntz, K., and Reiss, Jr., A. J. "Redisigning the Kansas City preventive patrol experiment", Evaluation 3 (1 and 2), (1972), 124-131.

[9] Finney, D. J. An Introduction to the Theory of Experimental Design, Chicago: The University of Chicago Press, 1960, Ch. 7.